

FINE TUNING CONSENSUS OPTIMIZATION FOR DISTRIBUTED RADIO INTERFEROMETRIC CALIBRATION

Sarod Yatawatta

ASTRON, The Netherlands Institute for Radio Astronomy,
The Netherlands.

Email: yatawatta@astron.nl

ABSTRACT

We recently proposed the use of consensus optimization as a viable and effective way to improve the quality of calibration of radio interferometric data. We showed that it is possible to obtain far more accurate calibration solutions and also to distribute the compute load across a network of computers by using this technique. A crucial aspect in any consensus optimization problem is the selection of the penalty parameter used in the alternating direction method of multipliers (ADMM) iterations. This affects the convergence speed as well as the accuracy. In this paper, we use the Hessian of the cost function used in calibration to appropriately select this penalty. We extend our results to a multi-directional calibration setting, where we propose to use a penalty scaled by the squared intensity of each direction.

Index Terms— Calibration, Interferometry: Radio interferometry

1. INTRODUCTION

Modern radio interferometric arrays deliver large volumes of data, in order to reach higher sensitivities yielding new science. To reach the full potential of such arrays, estimation of systematic errors in the data and correction for such errors (also called as calibration) is essential. This is not a trivial task for an array with hundreds of receivers that collect data over many hours and at thousands of different frequencies. A case in point being the square kilometre array (SKA), which is in the planning phase. Thus, there is an urgent need for computationally efficient and robust algorithms. On the other hand, there is a surge in research related to large scale and distributed data processing algorithms (also called as big-data), which we can exploit to solve some of these problems.

Our recent work [1] introduced distributed-calibration as a way of distributing the computational burden over a network of computers while at the same time improving the quality of calibration. We essentially exploited the continuity of sys-

tematic errors over frequency to enforce an additional constraint onto calibration. This reduces calibration to a consensus optimization [2] problem and we used alternating direction method of multipliers (ADMM) [3] as the underlying algorithm in the proposed distributed calibration scheme.

Consensus optimization, practically implemented with ADMM, has been extensively studied and is deployed in a wide variety of application areas (some recent examples are [4, 5, 6]). In addition, similar work is beginning to appear in radio astronomical imaging [7, 8, 9]. However, compared with other users of ADMM, we observe several unique properties of the calibration problem that we face. First, the cost function used in calibration is non-linear and non-convex. The systematic errors are mainly caused by directional effects such as the ionosphere and the receiver beam shape. Although we know the general properties of such errors, building an entirely accurate model (for instance for their variation with frequency) is not feasible. Hence, we enforce consensus only by using an approximate model, and this is clearly different and also more involved from most other applications. Indeed, other applications such as consensus averaging, where consensus is enforced on a constant value, use a perfect model. Furthermore, most other applications use complicated network topologies (that in turn affect the performance of ADMM) and on the other hand, in our case, we have a much simpler (and fully connected) network with one fusion center.

Of particular interest is the convergence rate of ADMM, which depends on many factors including the penalty parameter and the network topology [10]. In most cases, the penalty parameter is selected by trial and error, following some general guidelines [3]. However, for specific problems, better methods to select the penalty have been proposed [10, 11, 12]. Recent work [13] has suggested to select the penalty parameter as large as possible to make the objective function strongly convex. Hence for our problem, we study the Hessian of the cost function to select appropriate values for the penalty parameter. For calibration along multiple directions in the sky, we can select different penalty values along each direction. Intuitively, we select a large penalty along directions with

This work is supported by the European Research Council project LO-FARCORE, grant #339743.

higher signal where we have more confidence in our model. These directions are mostly close to the center of the field of view. On the other hand, for directions far away from the center, we select a smaller penalty.

The rest of the paper is organized as follows: In section 2 we give an overview of radio interferometric calibration. Next, in section 3, we present distributed calibration based on consensus optimization. We also present a scheme based on the Hessian of the cost function to select the penalty parameter. Simulation results are presented in section 4 where we demonstrate the improved performance with a refined penalty parameter. Finally, we draw our conclusions in section 5.

Notation: Matrices and vectors are denoted by bold upper and lower case letters as \mathbf{J} and \mathbf{v} , respectively. The transpose and the Hermitian transpose are given by $(\cdot)^T$ and $(\cdot)^H$. The matrix Frobenius norm is given by $\|\cdot\|$. The set of real and complex numbers are denoted by \mathbb{R} and \mathbb{C} . The identity matrix is given by \mathbf{I} . The matrix trace operator is given by $\text{trace}(\cdot)$.

2. RADIO INTERFEROMETRIC CALIBRATION

Consider a radio interferometric array with N receivers. The sky is composed of many discrete sources and we consider calibration along K directions in the sky. The observed data at a baseline formed by two receivers, p and q is given by [14]

$$\mathbf{V}_{pq} = \sum_{k=1}^K \mathbf{J}_{pk} \mathbf{C}_{pqk} \mathbf{J}_{qk}^H + \mathbf{N}_{pq} \quad (1)$$

where $\mathbf{V}_{pq} (\in \mathbb{C}^{2 \times 2})$ is the observed *visibility* matrix (or the cross correlations). The systematic errors that need to be calibrated for station p and q are given by the Jones matrices $\mathbf{J}_{pk}, \mathbf{J}_{qk} (\in \mathbb{C}^{2 \times 2})$, respectively. Note that since K directions are calibrated, for each station, there are K Jones matrices (so KN in total). The sky signal (or *coherency*) along the k -th direction is given by $\mathbf{C}_{pqk} (\in \mathbb{C}^{2 \times 2})$ and is known a priori. The values of $\mathbf{J}_{pk}, \mathbf{J}_{qk}$ and \mathbf{C}_{pqk} in (1) are implicitly dependent on sampling time and frequency of the observation. The noise matrix $\mathbf{N}_{pq} (\in \mathbb{C}^{2 \times 2})$ is assumed to have complex, zero mean, circular Gaussian elements.

Estimating the Jones matrices in (1) can be further simplified by using the space alternating generalized expectation maximization (SAGE) algorithm [15, 16]. In a nutshell, using SAGE algorithm, we can simplify calibration along K directions to K single direction calibration subproblems (see [16] for details). Calibration along the k -th direction is done by using the effective observed data

$$\mathbf{V}_{pqk} = \mathbf{V}_{pq} - \sum_{l=1, l \neq k}^K \hat{\mathbf{J}}_{pl} \mathbf{C}_{pql} \hat{\mathbf{J}}_{ql}^H \quad (2)$$

using current estimates $\hat{\mathbf{J}}_{pl}$ and $\hat{\mathbf{J}}_{ql}$ and for an array with N receivers, we can form at most $N(N-1)/2$ baselines that

collect visibilities as in (2), for any given time and frequency sample. We define our objective function (for the k -th direction) under a Gaussian noise model as

$$g_k(\mathbf{J}_{1k}, \mathbf{J}_{2k}, \dots) = \sum_{p,q} \|\mathbf{V}_{pqk} - \mathbf{J}_{pk} \mathbf{C}_{pqk} \mathbf{J}_{qk}^H\|^2 \quad (3)$$

where the summation is over the baselines pq that have data. By increasing the time and frequency interval within which data are collected, this summation can be expanded (thus improving the signal to noise ratio). By defining $\mathbf{J} (\in \mathbb{C}^{2N \times 2})$ as the augmented matrix of Jones matrices of all stations along the k -th direction,

$$\mathbf{J} \triangleq [\mathbf{J}_{1k}^T, \mathbf{J}_{2k}^T, \dots, \mathbf{J}_{Nk}^T]^T, \quad (4)$$

and $\mathbf{A}_p (\in \mathbb{R}^{2 \times 2N})$ (and \mathbf{A}_q likewise) as the canonical selection matrix

$$\mathbf{A}_p \triangleq [\mathbf{0}, \mathbf{0}, \dots, \mathbf{I}, \dots, \mathbf{0}], \quad (5)$$

(only the p -th block of (5) is an identity matrix) we can rewrite (3) as

$$g_k(\mathbf{J}) = \sum_{p,q} \|\mathbf{V}_{pqk} - \mathbf{A}_p \mathbf{J} \mathbf{C}_{pqk} (\mathbf{A}_q \mathbf{J})^H\|^2. \quad (6)$$

Calibration along the k -th direction is the estimation of \mathbf{J} by minimizing (6). Note that (6) has to be minimized for each direction $k = 1 \dots K$ and updated values of (2) are re-used until convergence is reached in the SAGE algorithm. We also note that (6) only gives solutions for one frequency and time interval, and to calibrate the full dataset, many such solutions are obtained for data observed at different time and frequency intervals.

3. DISTRIBUTED CALIBRATION

We have introduced calibration along K directions, but only working on a single frequency and time sample in section 2. In this section, we consider calibrating data observed at P different frequencies, but only along 1 direction, because this can easily be extended to K directions using the SAGE algorithm. We impose an additional constraint that tries to preserve continuity of \mathbf{J} in (6) over frequency. To solve this, we introduced the use of consensus optimization in [1], where the objective function is modified into an augmented Lagrangian

$$L_f(\mathbf{J}_f, \mathbf{Z}, \mathbf{Y}_f) = g_f(\mathbf{J}_f) + \|\mathbf{Y}_f^H (\mathbf{J}_f - \mathbf{B}_f \mathbf{Z})\| + \frac{\rho}{2} \|\mathbf{J}_f - \mathbf{B}_f \mathbf{Z}\|^2 \quad (7)$$

where the subscript $(\cdot)_f$ denotes data (and parameters) at frequency f . In (7), $g_f(\mathbf{J}_f)$ is the original cost function as in (6), except that the subscripts denote frequency f . The Lagrange multiplier is given by $\mathbf{Y}_f (\in \mathbb{C}^{2N \times 2})$. The calibration parameters are given by $\mathbf{J}_f (\in \mathbb{C}^{2N \times 2})$. The continuity in frequency is enforced by the frequency model given by $\mathbf{B}_f (\in \mathbb{R}^{2N \times 2NF})$, which is essentially a set of basis functions in

frequency, evaluated at f . The global variable $\mathbf{Z} (\in \mathbb{C}^{2NF \times 2})$ is shared by data at all P frequencies.

The ADMM iterations for solving (7) are given as

$$(\mathbf{J}_f)^{n+1} = \arg \min_{\mathbf{J}} L_f(\mathbf{J}, (\mathbf{Z})^n, (\mathbf{Y}_f)^n) \quad (8)$$

$$(\mathbf{Z})^{n+1} = \arg \min_{\mathbf{Z}} \sum_f L_f((\mathbf{J}_f)^{n+1}, \mathbf{Z}, (\mathbf{Y}_f)^n) \quad (9)$$

$$(\mathbf{Y}_f)^{n+1} = (\mathbf{Y}_f)^n + \rho((\mathbf{J}_f)^{n+1} - \mathbf{B}_f(\mathbf{Z})^{n+1}) \quad (10)$$

where we use the superscript $(\cdot)^n$ to denote the n -th iteration. The steps (8) and (10) are done for each f in parallel. The update of the global variable (9) is done at the fusion center. More details of these steps can be found in [1].

In this paper, we study strategies for selecting the penalty parameter ρ to get faster convergence and accurate results. In order to do this, we use the Hessian operator of the cost function (6), which is given as [1, 17],

$$\begin{aligned} \text{Hess}_f(g_f(\mathbf{J}), \mathbf{J}, \boldsymbol{\eta}) &= \sum_{p,q} (\mathbf{A}_p^T ((\mathbf{V}_{pqf} - \mathbf{A}_p \mathbf{J} \mathbf{C}_{pqf} \mathbf{J}^H \mathbf{A}_q^T) \mathbf{A}_q \boldsymbol{\eta} \\ &\quad - \mathbf{A}_p (\mathbf{J} \mathbf{C}_{pqf} \boldsymbol{\eta}^H + \boldsymbol{\eta} \mathbf{C}_{pqf} \mathbf{J}^H) \mathbf{A}_q^T \mathbf{A}_q \mathbf{J}) \mathbf{C}_{pqf}^H \\ &\quad + \mathbf{A}_q^T ((\mathbf{V}_{pqf} - \mathbf{A}_p \mathbf{J} \mathbf{C}_{pqf} \mathbf{J}^H \mathbf{A}_q^T)^H \mathbf{A}_p \boldsymbol{\eta} \\ &\quad - \mathbf{A}_q (\mathbf{J} \mathbf{C}_{pqf} \boldsymbol{\eta}^H + \boldsymbol{\eta} \mathbf{C}_{pqf} \mathbf{J}^H)^H \mathbf{A}_p^T \mathbf{A}_p \mathbf{J}) \mathbf{C}_{pqf}) \end{aligned} \quad (11)$$

where $\boldsymbol{\eta} \in \mathbb{C}^{2N \times 2}$.

For convexity, we need a positive definite Hessian. Since we have a Hessian operator (instead of a matrix), we need to find the smallest eigenvalue of the Hessian, and for convexity, this should be positive. In order to find this, we define a cost function as

$$\begin{aligned} h(\boldsymbol{\eta}) &\triangleq \frac{1}{2} \text{trace}(\boldsymbol{\eta}^H \text{Hess}_f(g_f(\mathbf{J}), \mathbf{J}, \boldsymbol{\eta}) \\ &\quad + \text{Hess}_f^H(g_f(\mathbf{J}), \mathbf{J}, \boldsymbol{\eta}) \boldsymbol{\eta}) \end{aligned} \quad (12)$$

and we find the smallest eigenvalue λ by solving

$$\begin{aligned} \lambda &= \arg \min_{\boldsymbol{\eta}} h(\boldsymbol{\eta}) \\ \text{subject to } &\boldsymbol{\eta}^H \boldsymbol{\eta} = \mathbf{I}. \end{aligned} \quad (13)$$

The constraint $\boldsymbol{\eta}^H \boldsymbol{\eta} = \mathbf{I}$ makes the minimization of (12) restricted onto a complex Stiefel manifold [18], which can be easily solved by using the Riemannian trust region method [19, 20]. In order to do this, we require the gradient and Hessian of $h(\boldsymbol{\eta})$, which are given as

$$\text{grad}(h(\boldsymbol{\eta}), \boldsymbol{\eta}) = \text{Hess}_f(g_f(\mathbf{J}), \mathbf{J}, \boldsymbol{\eta}) \quad (14)$$

and

$$\text{Hess}(h(\boldsymbol{\eta}), \boldsymbol{\eta}, \boldsymbol{\zeta}) = \text{Hess}_f(g_f(\mathbf{J}), \mathbf{J}, \boldsymbol{\zeta}), \quad (15)$$

where $\boldsymbol{\zeta} \in \mathbb{C}^{2N \times 2}$.

After obtaining λ from (13), our strategy is to select ρ such that $\rho + \lambda \geq 0$ so that the Hessian of the augmented Lagrangian (7) is positive semi-definite [13]. In order to do this, we need an estimate for \mathbf{J} in (12). We can find this by initial calibration with a pre-determined value of ρ (say $\rho = 0$). Once we obtain $\hat{\mathbf{J}}$, we use (13) to find λ and afterwards we update ρ . Note that λ is dependent on f , but we ignore the frequency dependence of λ and use one value of f (typically the middle) to estimate it.

So far, we have considered calibration along one direction only. The next question that we must answer is how to select ρ for calibration along K directions in the sky. For each direction, \mathbf{C}_{pqf} in (11) will influence the value of λ . If the centroid of the source (cluster) [21] is along l, m direction in the sky and if its effective (unpolarized) intensity is α , we have

$$\mathbf{C}_{pqf} \approx \exp(j\phi(l, m, p, q)) \alpha \mathbf{I} \quad (16)$$

where $\phi(l, m, p, q)$ is the phase contribution and \mathbf{I} is a 2×2 identity matrix. Hence \mathbf{C}_{pqf} is a diagonal scalar matrix. If $\hat{\mathbf{J}}$ is close to the true solution, the term $\mathbf{V}_{pqf} - \mathbf{A}_p \mathbf{J} \mathbf{C}_{pqf} \mathbf{J}^H \mathbf{A}_q^T$ becomes negligible compared with the other terms in (11). The remaining terms have a product $\mathbf{C}_{pqf} \mathbf{C}_{pqf}^H$ and the phase term in (16) cancel out. Therefore, for different clusters, the value for λ obtained by (13) is mainly determined by the squared effective intensity α^2 of each source. Hence, once we have determined a suitable value for ρ for one direction, the corresponding values for other directions can be determined by scaling by the squared effective intensity.

4. SIMULATION RESULTS

We simulate an array of $N = 47$ receivers that calibrate along $K = 5$ directions in the sky. The matrices $\mathbf{J}_{pk}, \mathbf{J}_{qk}$ in (1) are generated with their elements having values drawn from a complex uniform distribution in $[0, 1]$, multiplied by a frequency dependence given by a random 7-th order polynomial. The intensities of the $K = 5$ sources are randomly generated in the range $[1, 5]$ and their positions are randomly chosen in a field of view of about 7×7 square degrees. The variation of intensities with frequency is given by a power law with randomly generated exponent in $[-1, 1]$. The noise matrices \mathbf{N}_{pq} in (1) are simulated to have complex circular Gaussian random variables. The variance of the noise is changed according to the signal to noise ratio (SNR = 10)

$$\text{SNR} \triangleq \frac{\sum_{p,q} \|\mathbf{V}_{pq}\|^2}{\sum_{p,q} \|\mathbf{N}_{pq}\|^2}. \quad (17)$$

With this setup, we generate data for $P = 8$ frequency channels in the range 115 to 185 MHz. For calibration, we setup a 3-rd order polynomial model ($F = 4$), using Bernstein basis functions [22] for the matrix \mathbf{B}_f in (7). Note that we intentionally use a lower order frequency dependence than what is

actually present in the data to create a realistic scenario when the exact model is not known. During calibration, initial values for the parameters are always set as $\mathbf{J}_p = \mathbf{I}$ for $p \in [1, N]$. Unless stated otherwise, all directions have the same value of ρ . We use 50 ADMM iterations, and after the 1-st iteration, we solve (13) to estimate λ , and we get a typical value of $\lambda = -150$ for a source with unit amplitude. Regardless, we perform calibration with various values of ρ to compare performance.

We find the normalized (averaged over all directions) mean squared error (NMSE) between true \mathbf{J}_f and its estimate as

$$\text{NMSE} \triangleq \frac{1}{\sqrt{2KN}} \sqrt{\sum_k \|\mathbf{J}_f - \hat{\mathbf{J}}_f \mathbf{U}\|^2} \quad (18)$$

to measure the accuracy of calibration. In (18), \mathbf{U} is a unitary matrix that removes the unitary ambiguity in the estimated $\hat{\mathbf{J}}_f$ [23].

In Fig. 1, we show the NMSE for various values of ρ , with increasing number of ADMM iterations. We see that for $\rho + \lambda > 0$ ($\rho = 200$) we get the best performance, but increasing ρ too much beyond this value ($\rho = 1000$) shows no additional improvement. A notable behavior of the NMSE is the enhancement of the error at the edges (especially at low ADMM iterations), which we attribute to Runge's phenomenon [24] in polynomial interpolation. In Fig. 2, we show

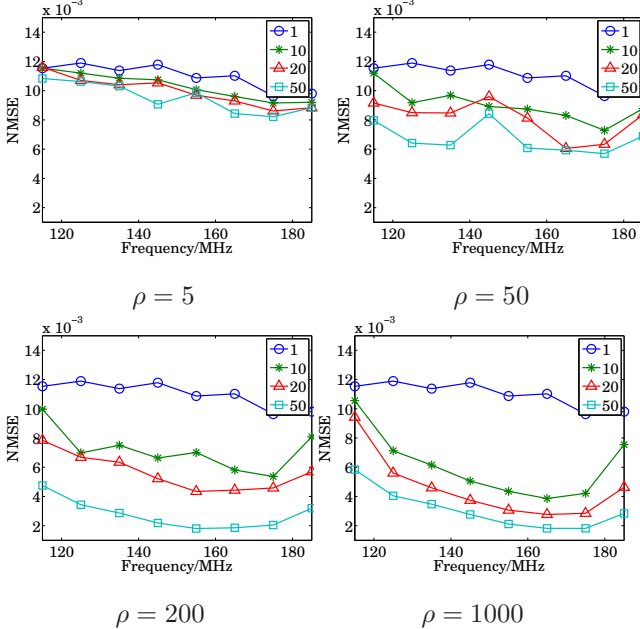


Fig. 1. NMSE for various ρ with increasing ADMM iterations.

the final NMSE for 50 ADMM iterations, which once again shows that $\rho = 200$ gives the best result.

In Fig. 3, we show NMSE for a simulation with intensities at mid frequency 5, 3, 3, 2 and 1.5 along the $K = 5$ directions.

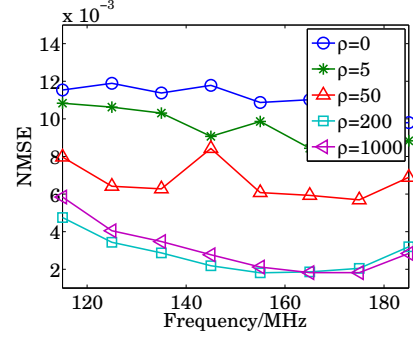


Fig. 2. NMSE for various ρ after 50 ADMM iterations.

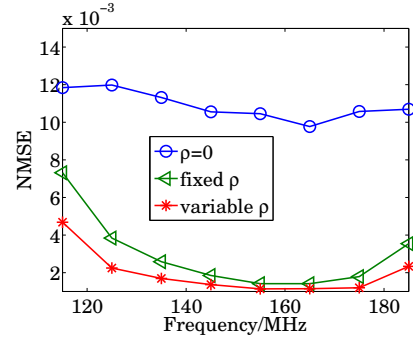


Fig. 3. NMSE after 50 ADMM iterations with fixed ρ along all directions and varying ρ according to squared intensity.

In one calibration, we use regularization $\rho = 400$ for all directions and in the other, we use ρ equal to 400, 144, 144, 64 and 36 respectively. We see that varying ρ in proportion to the squared intensity gives the better NMSE.

5. CONCLUSIONS

We have investigated refining the performance of distributed calibration based on consensus optimization in this paper. We have used the Hessian of the cost function to appropriately select the penalty parameter such that the augmented Lagrangian becomes convex. Furthermore, in a multi-directional calibration scheme, we have proposed to scale the penalty parameter proportional to the squared intensity along each direction. According to our simulations, such fine-tuning of parameters gives superior performance in terms of accuracy and convergence of the distributed calibration scheme.

6. REFERENCES

- [1] S. Yatawatta, "Distributed radio interferometric calibration," *Monthly Notices of the Royal Astronomical Society*, vol. 449, no. 4, pp. 4506–4514, 2015.

- [2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [3] D. Bertsekas and J.N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, 2nd edition, Singapore: Athena Scientific, 1997.
- [4] T.-H. Chang, M. Hong, and X. Wang, "Multi-agent distributed optimization via inexact consensus ADMM," *Signal Processing, IEEE Transactions on*, vol. 63, no. 2, pp. 482–497, Jan. 2015.
- [5] E. Wei and A. Ozdaglar, "Distributed alternating direction method of multipliers," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, Dec 2012, pp. 5445–5450.
- [6] T. Erseghe, "A distributed and scalable processing method based upon ADMM," *Signal Processing Letters, IEEE*, vol. 19, no. 9, pp. 563–566, Sept 2012.
- [7] A. Ferrari, D. Mary, R. Flamary, and C. Richard, "Distributed image reconstruction for very large arrays in radio astronomy," in *Sensor Array and Multichannel Signal Processing Workshop (SAM), 2014 IEEE 8th*, June 2014, pp. 389–392.
- [8] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, "PURIFY: a new approach to radio-interferometric imaging," *Monthly Notices of the Royal Astronomical Society*, vol. 439, pp. 3591–3604, Apr. 2014.
- [9] A. Onose, R. E. Carrillo, A. Repetti, J. D. McEwen, J.-P. Thiran, J.-C. Pesquet, and Y. Wiaux, "Scalable splitting algorithms for big-data interferometric imaging in the SKA era," *ArXiv e-prints*, Jan. 2016.
- [10] R. Nishihara, L. Lessard, B. Recht, A. Packard, and M. I. Jordan, "A general analysis of the convergence of ADMM," in *International Conference on Machine Learning* 32, 2015.
- [11] A. Teixeira, E. Ghadimi, I. Shames, H. Sandberg, and M. Johansson, "The ADMM algorithm for distributed quadratic problems: Parameter selection and constraint preconditioning," *Signal Processing, IEEE Transactions on*, vol. 64, no. 2, pp. 290–305, Jan 2016.
- [12] E. Ghadimi, A. Teixeira, I. Shames, and M. Johansson, "Optimal parameter selection for the alternating direction method of multipliers (ADMM): Quadratic problems," *Automatic Control, IEEE Transactions on*, vol. 60, no. 3, pp. 644–658, March 2015.
- [13] M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, April 2015, pp. 3836–3840.
- [14] J. P. Hamaker, J. D. Bregman, and R. J. Sault, "Understanding radio polarimetry, paper I," *Astronomy and Astrophysics Supp.*, vol. 117, no. 137, pp. 96–109, 1996.
- [15] J.A. Fessler and A.O. Hero, "Space alternating generalized expectation maximization algorithm," *Signal Processing, IEEE Transactions on*, vol. 42, no. 10, pp. 2664–2677, Oct. 1994.
- [16] S. Kazemi, S. Yatawatta, S. Zaroubi, P. Labropoulos, A.G. de Bruyn, L. Koopmans, and J. Noordam, "Radio interferometric calibration using the SAGE algorithm," *Monthly Notices of the Royal Astronomical Society*, vol. 414, no. 2, pp. 1656–1666, June 2011.
- [17] S. Yatawatta, "Radio interferometric calibration using a Riemannian manifold," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013, pp. 3866–3870.
- [18] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton Univ. Press, Princeton NJ, 2008.
- [19] P.-A. Absil, C. G. Baker, and K. A. Gallivan, "Trust-region methods on Riemannian manifolds," *Found. Comput. Math.*, vol. 7, no. 3, pp. 303–330, July 2007.
- [20] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, "Manopt, a Matlab toolbox for optimization on manifolds," *Journal of Machine Learning Research*, vol. 15, pp. 1455–1459, 2014.
- [21] S. Kazemi, S. Yatawatta, and S. Zaroubi, "Clustered calibration: an improvement to radio interferometric direction-dependent self-calibration," *Monthly Notices of the Royal Astronomical Society*, vol. 430, no. 2, pp. 1457–1472, 2013.
- [22] R. T. Farouki and V. T. Rajan, "Algorithms for polynomials in Bernstein form," *Comput. Aided Geom. Des.*, vol. 5, no. 1, pp. 1–26, June 1988.
- [23] S. Yatawatta, "On the interpolation of calibration solutions obtained in radio interferometry," *Monthly Notices of the Royal Astronomical Society*, vol. 428, no. 1, pp. 828–833, Jan. 2013.
- [24] G. Dahlquist and A. Björck, *Numerical Methods in Scientific Computing, Volume I*, Society for Industrial and Applied Mathematics, 2008.